

A Language-First Governance Framework for ShelfHub

A Preprint Server for LLMs and Humanity

Draft for ShelfHub Governance

March 29, 2026

Abstract

This paper presents a language-first governance framework and build blueprint for ShelfHub. Rather than auditing a fully verified deployed state, it specifies the architecture being constructed: search, recent-upload visibility, category browsing, multilingual support, API documentation, auto-review tooling, and an open submission path. The central claim is that, in a human–LLM writing environment, epistemic reliability cannot depend only on institutional filtering speed; it must also depend on disciplined language structures that are interpretable by both people and machines. The paper formalizes a five-part constructive method (Claim, Scope, Grounding, Uncertainty, Follow-up), introduces an entropy-to-order recovery model for open preprint ecosystems, and defines how API-mediated indexing may restore site-level coherence without centralized editorial gatekeeping. It concludes with bounded governance language that separates design intention from correctness certification while preserving transparency, interoperability, and iterative correction.

Keywords: preprints, human–LLM collaboration, open science, governance, epistemic reliability, machine-readable integration, constructive method, semantic recovery

1 Introduction

Human–LLM collaboration has changed how research writing starts, evolves, and circulates. Drafting speed has increased, iteration cycles have shortened, and the number of publicly shareable intermediate artifacts has grown. This acceleration expands participation, but it also introduces an epistemic governance problem: high-velocity publication can outpace traditional quality filters, especially when claims evolve through mixed human and model authorship. Empirical studies of preprint ecosystems indicate that rapid disclosure can improve circulation while simultaneously increasing heterogeneity in signaling and interpretation [1, 4].

Classical peer review remains important, but it is not designed to provide immediate triage for every fast-moving draft, replication note, or model-assisted revision. Its strengths—depth, selectivity, and field-specific scrutiny—can coexist with latency, access asymmetry, and limited throughput [16]. In parallel, the broader reproducibility debate has shown that institutional publication status alone does not guarantee durable correctness [6]. The relevant question for contemporary infrastructure is therefore not only “Is this claim true now?” but also “Can this claim be found, parsed, contested, and updated in a traceable way?”

A preprint server in this setting is not merely a file host; it is an epistemic coordination layer [18]. The practical challenge is to support rapid disclosure without overclaiming certification authority. This paper argues that one feasible path is language-first governance: requiring disciplined claim expression so that both human readers and LLM-based tools can perform

structured post-publication scrutiny. This requirement becomes more important when foundation models can generate fluent but potentially ungrounded text [3, 2, 7].

1.1 Research Question and Contribution

This paper asks: how can an open preprint platform preserve epistemic discipline without adopting centralized, journal-style gatekeeping?

It contributes four elements:

- a build-grounded governance framing derived from ShelfHub’s target architecture,
- a constructive five-part language protocol for claim discipline,
- an entropy-to-order recovery model explaining how LLMs can extract structured signals from noisy open submissions,
- a bounded policy language that separates discoverability from correctness certification.

1.2 Scope Note on Platform Status

This manuscript is a governance-and-design paper, not a deployment audit. As of March 29, 2026, statements about ShelfHub features in this document should be read as target architecture and active build intentions, unless explicitly marked as measured results. The next section therefore moves from design specification to normative governance proposals; ecosystem impacts are framed as hypotheses for later empirical testing.

2 ShelfHub Target Architecture Under Construction

In the current concept-to-implementation roadmap, ShelfHub is being built as more than a static upload page. The intended platform layers are:

- a planned top-level search interface for paper discovery,
- planned recent-upload visibility,
- planned category-based browsing,
- planned multilingual interface and content support,
- planned API documentation for machine access,
- a planned auto-review interface,
- a planned submission entry point.

Even as these modules are being implemented, this composition defines a clear product direction: discoverable, machine-aware, multilingual, and submission-driven scholarly communication. The governance proposal in this paper starts from this design target and evaluates its epistemic implications.

3 Theoretical Positioning

3.1 Open Science and FAIR Alignment

Open-science infrastructure emphasizes timely sharing, transparent reuse, and persistent accessibility. The FAIR principles (Findable, Accessible, Interoperable, Reusable) provide a direct vocabulary for evaluating whether research artifacts can travel across users, tools, and contexts [17]. In the current ShelfHub build plan, the search layer, API pathway, and multilingual interface are intended as practical FAIR enablers once implemented and documented. This orientation is consistent with reproducible-computation and open-research programs that argue for inspectable workflows as a condition of cumulative science [14, 12].

3.2 Commons Governance Perspective

Knowledge platforms also function as shared commons. Commons governance theory suggests that legitimacy can emerge from clear rules, visible responsibilities, and iterative monitoring rather than from one central authority alone [13]. Applied to preprints, this implies that explicit posting norms, transparent limitations, and update traceability can support trust even under open submission conditions.

3.3 Scientific Method, Falsifiability, and Error-Correction

From a natural-science perspective, governance language should privilege testable claims, explicit failure conditions, and iterative correction. This aligns with falsifiability-oriented views of scientific progress [15] and with program-level appraisal models in which competing explanations are compared over time through anomaly handling and progressive problem shifts [9]. It also recognizes that field-level consensus can reorganize when dominant frameworks accumulate unresolved tension [8]. In practical terms, platform governance should maximize opportunities for transparent error detection rather than assume that initial publication status confers stability.

3.4 Semantics and Generative Grammar as Disciplines

Language-first governance treats semantics as a continuity discipline and generative grammar as a structure discipline. Semantic continuity requires stable term usage or explicit redefinition; otherwise, retrieval systems and readers may infer false agreement. Structural discipline requires claims to appear in interpretable forms that preserve logical roles across revisions. This framing is consistent with long-running work on formal meaning and compositional structure [20, 19]. It also supports the social-epistemic view that objectivity is strengthened by publicly organized criticism across communities [10].

3.5 Philosophical Deepening: Language, Responsibility, and Epistemic Justice

From philosophy of language, research texts are not only containers of propositions; they are also performative commitments that position authors relative to evidence and community norms [22, 23]. In this view, C1–C5 functions as an illocutionary transparency protocol: it specifies

what kind of act is being performed (assertion, delimitation, justification, uncertainty disclosure, revision commitment), making later critique more precise.

A second point concerns meaning and rule-following. If meaning is stabilized through public use rather than private intention, term discipline and update traceability become epistemic necessities, not editorial preferences [24]. The governance objective is therefore not to freeze vocabulary, but to make semantic shifts explicit enough for shared interpretation across human readers and LLM pipelines.

A third point concerns fairness in knowledge participation. Open submission can broaden access, yet without explicit reasoning structure it can reproduce testimonial asymmetries in which some contributions are systematically discounted or misread. Language-first constraints can mitigate this by requiring reasons to be publicly inspectable and contestable, an orientation aligned with epistemic justice and communicative accountability [25, 26].

4 From Information Entropy to Order Recovery

4.1 Entropy in Open Submission Ecosystems

Open posting increases epistemic throughput but also increases heterogeneity in terminology, evidentiary quality, and argumentative completeness. In this paper, that condition is described as an information-entropy problem in an operational (not thermodynamic) sense: the local unpredictability of claim form rises as submission volume and stylistic variance rise [21]. This does not imply that open systems are inherently unreliable; it implies that reliability depends on whether communicative regularities are explicit enough to support downstream error-correction.

4.2 LLM-Based Semantic Recovery via Language Discipline

Entropy does not imply collapse if submissions expose regular linguistic anchors. The proposed constructive method (C1–C5) creates those anchors. When authors explicitly state claim type, scope boundary, grounding basis, uncertainty, and update linkage, LLM systems can recover structured signals from noisy corpora with lower ambiguity.

In this sense, language discipline acts as semantic scaffolding: it does not guarantee truth, but it makes comparison, contradiction detection, and synthesis computationally tractable. Given known risks of hallucination and ungrounded generation in language models, this scaffolding is a methodological requirement rather than a stylistic preference [7, 2].

4.3 Dynamic Index Formation Through APIs

Given machine-accessible metadata and full-text endpoints, post-publication models can iteratively construct relational indices across submissions. At minimum, they can label:

- **support relations** (one submission strengthens another claim),
- **conflict relations** (claims are mutually incompatible within overlapping scope),
- **extension relations** (new scope or method extends prior statements),
- **revision relations** (author-provided updates supersede earlier versions).

This process enables site-level order recovery without assuming a single central adjudicator. Governance shifts from ex-ante exclusion to ex-post structuring under transparent linguistic constraints.

5 A Verbal Constructive Method for Authors

Because platform tooling cannot carry all epistemic burden, rigor must remain author-centered and language-centered. Each major claim should include:

- C1. **Claim statement:** what is being asserted.
- C2. **Scope statement:** where the claim is intended to hold.
- C3. **Grounding statement:** why the claim is plausible.
- C4. **Uncertainty statement:** what remains unresolved.
- C5. **Follow-up note:** how updates, corrections, or extensions will be linked.

5.1 Technical Significance for Machine-Readable Integration

The C1–C5 structure is not only rhetorical guidance; it is a parseable schema for mixed human–machine review.

- C1 creates a proposition node that can be indexed and compared.
- C2 provides boundary conditions that reduce false contradiction flags.
- C3 exposes evidentiary or inferential support for downstream audit.
- C4 encodes epistemic status, preventing overconfident aggregation.
- C5 creates temporal links required for version-aware synthesis.

A minimal machine-readable representation can therefore associate each submission with a tuple $s_i = (C1_i, C2_i, C3_i, C4_i, C5_i)$, plus cross-document edges generated from API-level analysis. This does not replace expert evaluation, but it improves triage quality, retrieval precision, and update traceability at scale. Comparable reproducibility-focused workflows suggest that lightweight structural constraints can significantly improve later-stage verification and reuse [14, 12].

5.2 Natural-Scientific Testability

To keep the framework scientifically meaningful, its effects should be evaluated as testable hypotheses rather than normative assumptions. A minimal evaluation program can specify:

- H1. **Extraction hypothesis:** C1–C5-compliant submissions increase inter-rater agreement and model agreement when extracting core claims.
- H2. **Contradiction hypothesis:** explicit scope statements reduce false contradiction detections in cross-paper relation mining.

H3. **Update hypothesis:** explicit follow-up notes improve recall of version links and correction trajectories.

These hypotheses can be tested with blinded annotation studies, precision/recall analysis of relation extraction, and calibration-oriented scoring of uncertainty quality [5]. This moves governance discussion from abstract preference to measurable performance.

6 Governance Language for the ShelfHub Build Program

Given ShelfHub’s target architecture, governance can remain modest and still be strong.

6.1 Core Principles

- G1. **Open submission:** lawful and policy-compliant work should be postable, subject to posted moderation and legal policies.
- G2. **Non-endorsement:** publication does not imply correctness certification.
- G3. **Human accountability:** authors remain responsible for claims, including model-assisted sections.
- G4. **Interface transparency:** platform features and limitations are stated plainly.
- G5. **Human + machine access:** design decisions support both reader comprehension and API utility.
- G6. **Multilingual accessibility:** core discovery and submission pathways remain usable across languages.

6.2 Minimum Submission Expectations

- disclose substantive model assistance,
- link major claims to evidence or explicit argument sections,
- state meaningful limitations,
- provide a visible path for updates or addenda,
- keep key claim terms stable or explicitly redefined across versions.

7 Discussion: Honest Limits and Reliability

A healthy governance posture requires explicit limits.

- Posting is not equivalent to correctness certification.
- Auto-review support is not a replacement for field-level expert judgment.
- Visibility is not the same as validation.
- API accessibility does not guarantee semantic quality by itself.

These limits increase, rather than reduce, platform credibility because they align claims with operational reality. A system that clearly states what it does not certify is less likely to produce false authority effects.

Philosophically, this non-endorsement principle can be read as institutionalized epistemic humility: the platform is neither a truth-maker nor a neutral void, but a structured arena for reason-giving and contestation. Reliability then emerges from dialogical testing of claims over time, rather than from preemptive authority signals [26, 10].

Several failure modes also remain: performative template compliance without substantive rigor, strategic ambiguity in scope statements, and multilingual drift in key terms. For this reason, language-first governance should be treated as a necessary but insufficient layer that complements, rather than displaces, domain expertise and post-publication critique. This conclusion is consistent with social-epistemic accounts in which reliability emerges from organized criticism and transparent responsiveness to objections [10, 25].

8 Near-Term Improvement Path

Without pretending to be a journal system, the ShelfHub build program can progress through near-term milestones:

- normalize metadata fields across categories,
- publish a lightweight C1–C5 tagging guide for contributors,
- expose API fields for contradiction and complementarity annotations,
- standardize update-link practices for revised posts,
- make auto-review output explicitly advisory,
- publish confidence scales and error taxonomies for auto-review feedback,
- expose stable API examples for reproducible integrations,
- maintain multilingual quality control for metadata and contributor guidance.

These are incremental improvements that match the target architecture while directly supporting entropy-to-order recovery in practice.

9 Conclusion

ShelfHub’s significance does not require maximal claims. A preprint server that combines submission, discovery, API access, multilingual pathways, and advisory tooling can provide meaningful public infrastructure for the LLM era when these components are implemented and maintained.

The decisive challenge is not whether the platform can replace all legacy institutions. It is whether it can remain honest about limits while reliably supporting open, structured, and reusable knowledge flow under conditions of high informational entropy. Language-first governance addresses this challenge by treating disciplined claim expression as a shared protocol for human critique and LLM-assisted semantic recovery.

In this model, trust is not produced by institutional symbolism alone. It is produced by transparent limits, interoperable structures, and continuous post-publication organization. The next research phase is empirical: testing whether C1–C5 compliance measurably improves extraction quality, contradiction mapping, and update tracking across disciplines.

Appendix: One-Page Operating Statement

- A1. ShelfHub is a preprint infrastructure, not a correctness authority.
- A2. Posting, discovery, and multilingual access are target core public services in the build roadmap.
- A3. Open submission implies entropy; governance must therefore prioritize recoverable structure.
- A4. API access is a first-class path for machine interoperability and dynamic indexing.
- A5. Auto-review features are advisory, not determinative.
- A6. C1–C5 language discipline is the minimum semantic scaffold for mixed human–LLM evaluation.
- A7. Author clarity and declared uncertainty remain central quality norms.

References

- [1] Richard J. Abdill and Ran Blekhman, Tracking the popularity and outcomes of all bioRxiv preprints, *eLife*, 8:e45133, 2019.
- [2] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, Proceedings of FAccT, 2021.
- [3] Rishi Bommasani et al., On the Opportunities and Risks of Foundation Models, *arXiv:2108.07258*, 2021.
- [4] Nicholas Fraser et al., The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape, *PLoS Biology*, 19(4):e3000959, 2021.
- [5] Tilmann Gneiting and Adrian E. Raftery, Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, 102(477):359–378, 2007.
- [6] John P. A. Ioannidis, Why Most Published Research Findings Are False, *PLoS Medicine*, 2(8):e124, 2005.
- [7] Ziwei Ji et al., Survey of Hallucination in Natural Language Generation, *arXiv:2202.03629*, 2023.
- [8] Thomas S. Kuhn, *The Structure of Scientific Revolutions*, University of Chicago Press, 1962.
- [9] Imre Lakatos, *The Methodology of Scientific Research Programmes*, Cambridge University Press, 1978.

- [10] Helen E. Longino, *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton University Press, 1990.
- [11] Robert K. Merton, The Normative Structure of Science, in *The Sociology of Science*, 1942.
- [12] Brian A. Nosek et al., Promoting an open research culture, *Science*, 348(6242):1422–1425, 2015.
- [13] Elinor Ostrom, *Governing the Commons*, Cambridge University Press, 1990.
- [14] Roger D. Peng, Reproducible Research in Computational Science, *Science*, 334(6060):1226–1227, 2011.
- [15] Karl R. Popper, *The Logic of Scientific Discovery*, Hutchinson, 1959.
- [16] Richard Smith, Peer review: a flawed process at the heart of science and journals, *Journal of the Royal Society of Medicine*, 99(4):178–182, 2006.
- [17] Mark D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3:160018, 2016.
- [18] Paul Ginsparg, ArXiv at 20, *Nature*, 476(7359), 2011.
- [19] Noam Chomsky, *Syntactic Structures*, Mouton, 1957.
- [20] Richard Montague, *Universal Grammar*, *Theoria*, 36(3), 1970.
- [21] Claude E. Shannon, A Mathematical Theory of Communication, *Bell System Technical Journal*, 27(3):379–423, 1948.
- [22] J. L. Austin, *How to Do Things with Words*, Oxford University Press, 1962.
- [23] John R. Searle, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, 1969.
- [24] Ludwig Wittgenstein, *Philosophical Investigations*, Blackwell, 1953.
- [25] Miranda Fricker, *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford University Press, 2007.
- [26] Jurgen Habermas, *The Theory of Communicative Action, Volume 1*, Beacon Press, 1984.